**23 Feb 2015-02-23**

**by Afonso Duarte**

**ELIXIR-EGI pilot Project**

**Actions report**

**Task 5: Analysis tools to work with data replicas**

*Identify and provide guidance for the use of key life science software applications and tools that can be used to work with reference datasets on EGI. These tools can be used by life science researchers to define and execute custom analysis that work on reference datasets hosted on EGI. The task will review the identified tools on the distributed testbed of Task 4, and will provide information for the users about these tools at a central location, ideally as software profiles in EGI AppDB. → Deliverable 4 (Task 5 will provide information about key life science software applications and tools that can be used by life science researchers to define and execute custom analysis that work on reference datasets hosted on EGI. The information will be published at some central location, ideally as software profiles in EGI AppDB.)*

Initial action focused on listing the applications using the datasets and databases indicated in T1 and T2. This list is presented in Resources for reference dataset replication table (<https://docs.google.com/spreadsheets/d/1gcSNCNBdwIgNN8URrj0b2w4WT18VA04BHOi5csT0HGM/edit?usp=sharing>). It includes: the name of database/dataset, application name, application contact person and URL of application. (PS: this list is in constant update).

This list shows the existence of different types of applications (and this is widely known for this kind of applications):

1. web-based applications running the query or calculation and only using a small data set. Typically with low flexibility and for small data sets
2. A majority of these web-based apps offer the possibility to download a more extended and versatile version of the applications. These versions are more flexible; however some types of platforms (e.g. Java) limit their use.
3. Command line application. For download and to be run in own computer. Fully flexible.

Besides these types one can identify several database independent applications, i.e. from as external developer or user community. These applications typically also bridge to specific communities and niches of research.